Predicting Football Player Performance for NU Football Using an Athletic Mindset Indicator

Ada Ng <adang@u.northwestern.edu>
Santhosh Subramanian <santhoshsubramanian2017@u.northwestern.edu>
Shanil Sharma <ShanilSharma2017@u.northwestern.edu>
Siddharth Muthukumaran <siddharthmuthukumaran2018@u.northwestern.edu>

EECS 349: Introduction to Machine Learning
Spring 2017
Professor Doug Downey
Northwestern University

INTRODUCTION

Our team collaborated with Northwestern University Football with the goal of improving varsity team performance by making predictions using data from a variety of sources including wearable biometric sensors, game-day play analysis, and mindset indicator surveys. After reviewing available data, we decided to focus on predicting the overall season ‘player grade’ for an individual if that candidate were to join the NU Football team based on a psychometric assessment survey.

ProFootball player grades are assigned by a human grader (a former coach or other football expert) after watching the video of a game and consists of assigning each player a score from -2 to +2 in 0.1 increments based on their performance. An individual’s player grades are added together for a season player grade; the higher the grade, the more valuable the player. The Northwestern University men’s football team brought in $36,212,907 between the years 2015-2016 (Source: US Department of Education Equity in Athletics Data Analysis. https://ope.ed.gov/athletics/#/institution/details. Accessed April 19, 2017). The ability to predict future player performance at time of recruitment can allow NU Football to spend scholarship and other recruitment funds with greater return on investment. Additionally, improved team composition and performance can lead to increased revenue for the football program.

We hypothesize that the features of the mindset indicator survey are significantly correlated with the player grade.

DATA ACQUISITION AND PREPROCESSING

Each member of the NU Football team is administered an annual psychometric assessment survey (Troutwine Athletic Profile (TAP) score) (http://therightprofile.com/), the result which is a score comprised of 21 attributes. These attributes are all numeric with a mix of scores out of 100 or on a scale of 1-5. Examples of TAP scores include:

- Super Mental Toughness (1-100)
- Combat Score (1-100)
- Confidence (1-100)
- Desire (1-5)
- Coachability (1-5)

We received the TAP scores for each player for games from 2013 through 2016 - a total of 434 samples.

ProFootball Player Grades are assigned by a human grader (a former coach or other football expert) after watching the video of a game and consists of giving the player a score from -2 to +2 in 0.1 increments based on their performance. An individual’s player grades are added together for an overall season player grade.

We received season player grades for each player for games from 2014 to 2016 - a total of 338 samples.

We merged the two data sets by manually matching a player’s name. A sample was created when a TAP score and an season player grade for a single player was found at maximum one year apart. That is, if a teammate took a TAP survey in 2016, but the only season player grade available for this teammate was in 2014, we would discard the data. This resulted in far fewer data points because there were also players that had TAP scores on record but no annual player grade and vice versa. An individual player was allowed to be reflected in our final data set multiple times if they had multiple TAP scores and annual player grades that would match in multiple years as we operated under the assumption that an individual’s TAP results and performance can vary year by year.

Ultimately, we ended up with 73 samples with a season player grade range of -30.6 to 38.

METHODS and RESULTS

We experimented with the data using several different data transformations and learners in an effort to reach a minimal accuracy of 80%; each of these methods are outlined in sections below. For each method below, we employed cross validation with 80% of the data used for training and 20% used for testing.

Regression with total annual player grade

Our feature set consists of numerical attributes and our desired output is also continuous. For this reason, we initially experimented using several different regression models starting with linear regression. The output was a rather high root mean square error (RMSE) (See Table 1) so we looked into reducing the feature space first looking at step-wise to identify the strongest feature space as it ignores irrelevant features.

When step-wise with linear regression did not help to reduce RMSE, we employed ridge regression and lasso regression which are forms of regularized regression and include parameter shrinkage and variable selection.

Table 1. Regression RMSE Results for Season Player Grade

Model	**Root Mean Square Error (RMSE)**
Linear Regression	8.11386
Linear Regression (using only features from step-wise)	8.10873
Ridge Regression	8.11395
Ridge Regression (using only features from step-wise)	8.10892
Lasso Regression	8.11231
Lasso Regression (using only features from step-wise)	8.11132

Considering that season player grades ranged from -30.6 to 38, this meant our model was not producing significant accurate predictions.

Figure 1. Heat Map of Correlation Metrics (season player grade, continuous output)

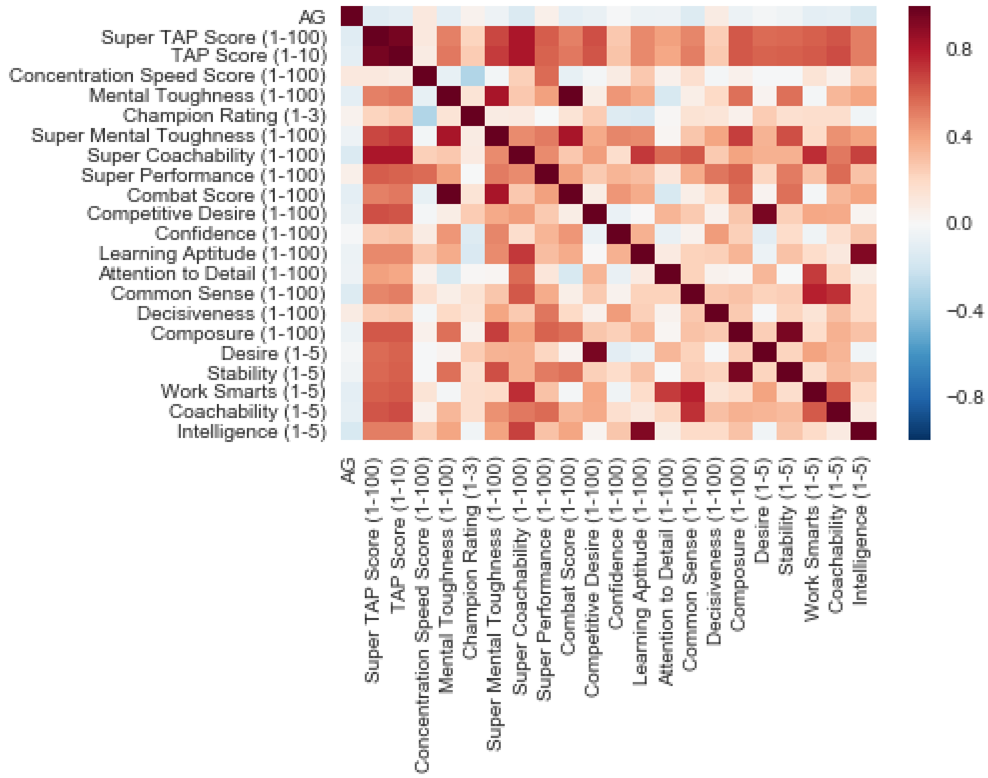
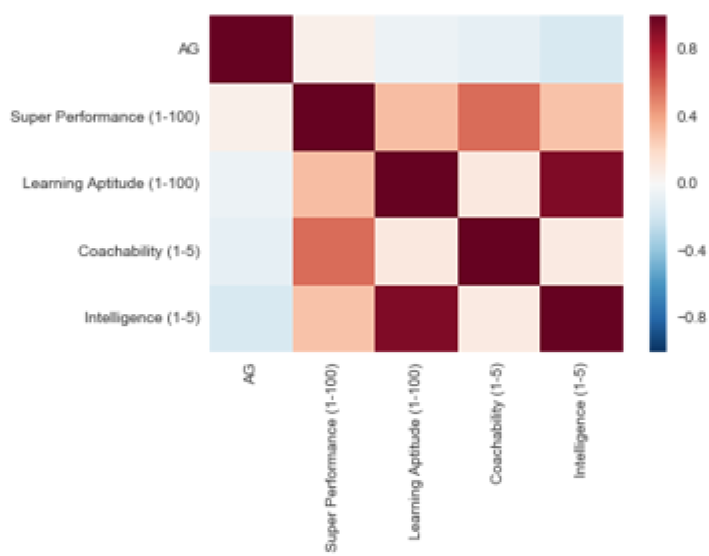


Figure 2. Heat Map of Correlation Metrics for Step-wise Attributes (season player grade, continuous output)



Discretizing Output
We decided to narrow down the output space to a simple binary positive season player grade or negative season player grade with the idea that it could help increase our accuracy. We used learners that are known to work well with categorical outputs with both the full feature space and a reduced feature space based on a step-wise function which determined that the strongest attributes were Super Performance, Learning Aptitude, Coachability, Intelligence.

Table 2. Classification Accuracy for Season Player Grade

Model	**Accuracy (using ALL features)**	**Accuracy (using 4 step-wise features)**
SVM	67%	53%
Logistic Regression	53%	60%
Naive Bayes	45%	47%
Nearest Neighbor	66%	29%
Random Forest	33%	46%

Figure 3. Heat Map of Correlation Metrics (season player grade, binary output)

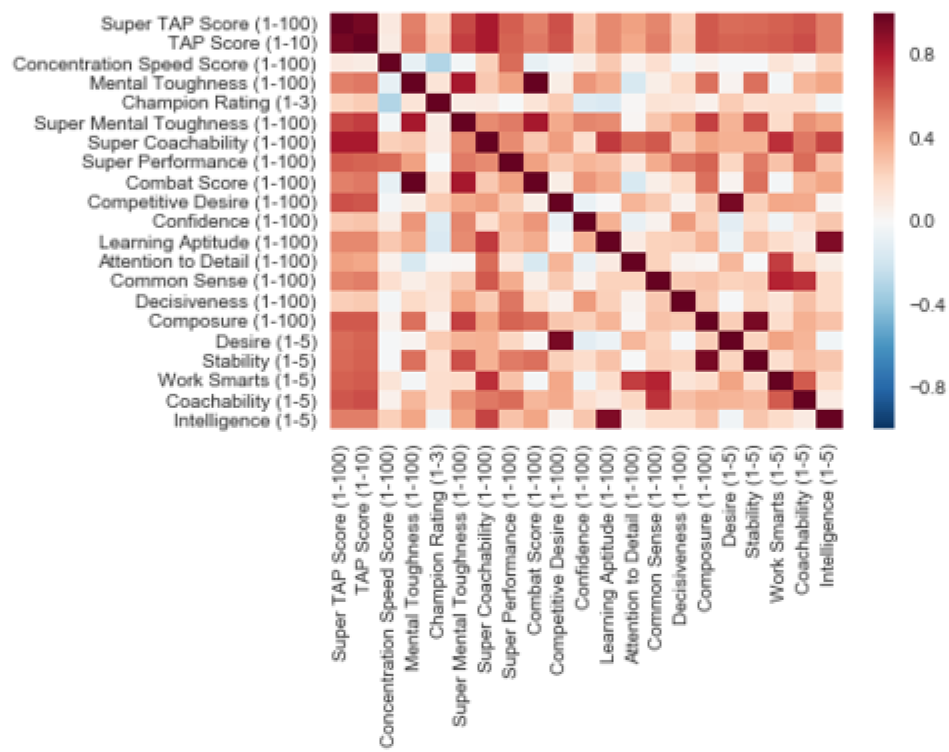


Figure 4. Heat Map of Correlation Metrics for Step-wise Attributes (season player grade, binary output)

FINDINGS

The highest accuracy we were able to achieve was using SVM and all features. However, we did not reach our accuracy goal of 80% and our p-value was 0.6929 which is well above the commonly accepted p-value of 0.05.

Hence, we don't have sufficient evidence to reject the null hypothesis; none of the features of the TAP test have been found to be significantly predictive of a player's grade per game.

LIMITATIONS AND FUTURE WORK

Our results may reflect a variety of factors from low reliability due to a low sample size or possibly a weak or no correlation between TAP score and player performance.

Future work can include repeating our experiments with a data size of at a minimum 500 samples. Further work could also segregate data between players of different positions. Perhaps there is better correlation between TAP score and player performance for certain positions than others.

MEMBER CONTRIBUTIONS

Ada Ng <adang@u.northwestern.edu> data collection, data preprocessing, result analysis, website content and project report

Santhosh Subramanian <santhoshsubramanian2017@u.northwestern.edu> data collection, web site, feature selection, model generation and result analysis

Shanil Sharma <ShanilSharma2017@u.northwestern.edu> data collection, web site, feature selection, model generation and result analysis

Siddharth Muthukumaran <siddharthmuthukumaran2018@u.northwestern.edu> data collection, data preprocessing, feature selection, model generation and result analysis